OTISS, A.S.

An Absolute Point Scale
for the Group Measurement
of Intelligence

Digitized by the Internet Archive
in 2014

https://archive.org/details/absolutepointsca00otis

Psych.
O88

150.117
150.1

An Absolute Point Scale for the Group Measurement of Intelligence

BY
ARTHUR S. OTIS
Surgeon General's Office, Washington, D. C.

UNIVERSITY OF TORONTO LIBRARY

1525661
25/8/20

(Reprint from THE JOURNAL OF EDUCATIONAL PSYCHOLOGY, Vol. IX, Nos. 5, 6, May-June 1918)

AN ABSOLUTE POINT SCALE FOR THE GROUP MEASUREMENT OF INTELLIGENCE.

ARTHUR S. OTIS

Surgeon General's Office, Washington, D. C.

CONTENTS

*The writer is indebted to Dr. Lewis M. Terman of Stanford University for many helpful suggestions during the making of this study.

I. Introduction

Purpose—The purposes of this study are:

(1) To construct a scale for the measurement of general mental ability, such scale being:

(a) suitable primarily for administration to the pupils in grades 4, 5, 6, 7, and 8 of the elementary school,

(b) capable of being administered to *groups* of at least 50,

(c) so constructed that the scoring is both rapid and, as far as possible, free from the error of the personal factor,

(d) built upon the general plan for an Absolute Point Scale outlined by the writer (see Ref. 10). Such construction would involve the "validation" and "graduation" of the tests, the determination of the probable error of a determined measure of general mental ability, etc.

(2) To investigate the correlation of the mental abilities tested by the scale.

It is not deemed feasible, in the space to which this article is limited, to attempt to discuss the nature of intelligence, such as it is presumed to measure by the present scale, nor the various definitions of intelligence which have been given or may be implied by the various 'intelligence scales' in present use. These subjects will be touched upon in various connections in the discussion. The writer will therefore proceed immediately to describe the manner in which the present scale was constructed.

II. The Tests

Requirements of a Scale for Mass Testing.—The chief object of testing in groups, of course, is economy of time. One of the most essential means of accomplishing this purpose is that the responses required be very simple. This makes for speed both in the administration and in the scoring of the tests. It has been the aim, therefore, so to arrange the tests in the present scale that there would be only one correct answer to each item and that this might be indicated merely by making a letter or figure or drawing a line. Where convenient, provision was made for the responses to be placed in a single column in which case the papers may be scored with dispatch

by the use of scoring forms. When every answer is either right or wrong, a large amount of time is saved that might be necessary otherwise to determine the value of partially correct answers. Moreover, under these conditions the tests must be scored in the same way by all investigators, this assuring comparability.

The ideas for the tests have been derived from various sources, chiefly, perhaps, from the Stanford Revision of the Binet Scale. (see Ref. 14). The general character of most of them is no doubt familiar.* In substance the remaining tests were designed especially for this study.

Description of the Tests.—The scale was compiled in duplicate. There were, in other words, two complete tests of each kind. The two tests of each kind were made as nearly alike as possible without using the same material. In each scale the tests were constructed as follows:

The Spelling Test† consisted of fifty pairs of words in two columns. The words of each pair consisted of the correct and incorrect spelling of a single word. In some cases the first spelling was the correct one and in some cases the second was the correct one. The pupil was required to indicate by the letters, F, S, or N, placed in a parenthesis opposite the words, as shown in the sample in Appendix I, whether the first or the second was the correct spelling, nor neither spelling was correct.

The Arithmetic Test consisted of 16 problems in which the computation was made as easy as possible and the emphasis thus placed upon reasoning.

The Synonym and Antonym Test consisted of 50 pairs of words as shown in the sample. The pupils were required to indicate by the letters, S and O, whether the words of a pair meant the same or the opposite.

The Proverb Test consisted of 20 proverbs in two sets of ten proverbs, each set followed by twelve statements, one of which "explained" each of the ten proverbs, there being two extra statements in each set not explaining the proverbs. The pupils were required to place in the parenthesis before each proverb the number of the statement which explained it.

*Thanks are due to Mrs. Mary D. Chamberlain for the Proverb Test used in this study. The words of the Spelling Test, as explained later, were taken from Ayres' list. (See Ref. 2.)

†Although the Spelling Test was found, in both the preliminary and in the present investigations, to afford quite as good a measure of intelligence, from one point of view, as the other tests (see intercorrelations), it has since been considered best to drop this test from the scale.

The Disarranged Sentence Test consisted of 26 sentences with the words disarranged, as shown in the sample. The pupils were required to rearrange the words mentally to make sense and indicate whether the sentences so constructed were true or false by underlining the words true or false at the end of the line.

The Relation Test consisted of 24 items, each in the form of a proportion in which one of the four terms was to be supplied, indicating by number from five alternative answers given on the same line.

The Geometric Test consisted of 22 items, using as a basic principle that described by Abelson. (See Ref. 1.) Referring to the figures constructed by overlapping one or more circles, triangles, and rectangles, the pupils were required to place figures 1, 2, etc., in certain designated spaces as suggested in the sample.

The Following Directions Test consisted of 14 problems requiring the pupils to place certain numbers in certain figures on the Woodworth and Wells Cancellation Sheet. This test presumably differed from the preceding one in that its difficulty consisted more in the comprehension of involved language while in the Geometric Test, the difficulty lay chiefly in tracing out the space relations.

The Narrative Completion Test was of the type used by Whipple, Ebbinghaus, Terman, and others. It consisted of a short story of which certain words were omitted leaving blanks in which the pupils were required to write the words which in their judgment best fitted into the story.

III. The Preliminary Investigation

To aid in choosing the tests for the Point Scale and in determining the most suitable forms in which to give them, a preliminary investigation was conducted in which 29 pupils in grades 4 to 8 of a small school were tested. Fifteen tests were used. Eight of these were in the same or nearly the same form as those shown in the Appendix. Two others, the Arithmetic and Spelling Tests, have since been entirely made over. The Synonyms Test was given orally first, then about three weeks later repeated in the form shown. The other five tests were: a test in word meaning recognition, of the type suggested by the writer (See Ref. 8) and called the Reading Test; a test in the reproduction, in writing, of sentences dictated, called the Memory for Sentences Test; the Trabue Completion Test (see Ref. 15); the Kansas Silent Reading Test (see Ref. 4); and the Starch Grammar Test (see Ref. 12). The list is shown in Table I. The tests marked with the asterisk were given in duplicate, the double scores being used in the correlations.

The correlation of each test with the composite of all scores except those of the Oral Synonyms and Grammar Tests are shown in the first column of the table. The correlations with mental age as determined by the Stanford Revision of the Binet Scale are shown in the second column. Considering the small number of individuals as well as the unreliability of scores, the coefficients may be regarded as of suggestive value only.

TABLE I.

Some Results of the Preliminary Investigation

Test	Correlation with Composite	Correlation with Mental Age
Relation	.94	.97
Proverbs*	.94	.94
Following Directions	.86	.95
Geometric*	.89	.92
Trabue Completion Test*	.88	.88
Reading*	.92	.82
Kansas Silent Reading Test	.90	.88
Synonyms (Oral)	.79	.87
Synonyms (Written)	.83	.85
Disarranged Sentences*	.86	.81
Narrative Completion	.86	.80
Arithmetic	.84	.80
Spelling*	.79	.84
Memory for Sentences	.77	.82
Memory for Digits*	.72	.42
Starch Grammar Test	.30	.49
Correlation of Mental Age with Composite Score: .94		
Same, "corrected for attenuation" (estimate): .99		

IV. Acquisition of the Data

Administration of the Tests.—The tests were given to 121 children of a large grammar school—43 in the fourth grade, 40 in the sixth grade, and 38 in the eighth grade. Each test was taken by all of the pupils of one grade at a time in their regular room, the teacher being present. The writer personally conducted all the tests, giving all the directions and explanations. The tests were given in approximately the order indicated in the foregoing section. In most instances two test series were given to each grade each day, one in the morning and one in the early afternoon. The giving of the first and second tests of the same kind were separated by three or more days in most instances but particularly in the cases of the Arithmetic, Geometric, and Following Directions Tests, in

*Double scores used.

which the second test is in the nature of a recast of the first. Time was allowed for all to finish in nearly all cases, except, of course, the Disarranged Sentence Test, which is a speed test. Occasionally when one or two pupils lagged far behind the others, their papers were taken up before they finished. In such cases it was usually noted that the pupils had permitted themselves to be distracted from their work. In the Disarranged Sentence Test sufficient time was allowed for only one pupil to finish. The order, "Stop," was then given and the time noted. For purposes of comparison, all scores were afterward increased to a five minute basis.

The pupils of each grade were adjured at the beginning of the testing not to give or receive aid during the taking of any tests. A wholesome attitude appeared to be taken by all during the testing. In such instances of apparent collusion as were noted, the pupils were quietly cautioned. These instances were few. On the whole the pupils were orderly and attentive and signified their interest in the testing.

Scoring.—In the case of each test except the Synonyms, Spelling, and Disarranged Sentences, one count was given for each correct answer and no count for incorrect or omitted answers. In the case of Synonyms, however, since there are but two alternative answers, S or O, theoretically, of the answers given concerning the pairs of words not known by any pupil, but guessed at, one half will be right by chance. Therefore, if say 35 of the 50 were known and correctly marked, and 10 of the remaining 15 guessed at, leaving 5 blank; of the 10 guessed at, 5 might be marked rightly by chance. This would make 40 correct, 5 incorrect, and 5 blank. It seemed, therefore, that as many counts should be deducted from the total correctly marked (40) as were incorrect (5) thus giving a score of $40 - 5 = 35$, the number assumed to be known. A person guessing at all of them and getting half right by chance would then attain a score of $25 - 25 = 0$. This method was adopted in scoring the Synonyms and Disarranged Sentence Tests. The case of the Disarranged Sentences is complicated by the fact that sentences wrongly marked on account of haste are penalized additionally by the loss of time in scanning them. The suitability of the method, therefore, should perhaps be investigated.

Since there are three possible answers, F, S, or N, in the Spelling Test, theoretically $\frac{1}{3}$ of the number of those guessed at would be marked rightly by chance. This would mean that, to follow the

above method, the score should be obtained by deducting from the number rightly marked, ½ the number of those wrongly marked. However, inasmuch as there were only fourteen individuals who did not attempt all the words, and to avoid possible negative scores, the scores were obtained by giving one count for each right answer, no count for each wrong answer, and ⅓ count for each blank, on the assumption that if guessed at, ⅓ of these words would have been rightly marked, This brought all the scores to the same basis and necessitated counting only right answers in all but 14 cases. Identical rank orders of the individuals are obtained from the scores by the two methods. The scores that would have been obtained by using the first method can easily be derived from those used merely by multiplying by 1½ and subtracting 25.

In order to obtain a suggestion as to the value of the above methods of scoring, the sum of the differences between the first and second scores of the 14 pupils above mentioned was found first when the scores were obtained as above and second when obtained by merely counting the number of correct answers, taking no account, therefore, of the element of chance. The sum of the differences in the first case was 34 and in the second 45, although the scores were less in the second case. This suggests that the method employed was the more reliable.

While there is, of course, a 'one-in-five chance' of an element of the Relation Test being marked rightly by guess, it was not deemed necessary to take account of it. In an auxilliary investigation regarding the scoring of the Digit Test, the papers were scored (1) according to the number of digits in the last number correctly reproduced, (2) according to the number of digits in the next to the last number correctly reproduced, and (3) according to the last group of numbers of the same size of which two or more were correctly reproduced. No one of these three methods appeared to be appreciably superior to the others. The reliability coefficient of scores by method (1) was .53, by the method employed in this study, .74. It was discovered after giving the test that some of the pupils were able to reproduce numbers of nine or more digits within three trials. If such had been included in the test, the reliability coefficient by method (1) would no doubt be higher. It is believed, therefore, that with a sufficiently exhaustive test, the loss in reliability of method (1) would be more than made up by the great saving in time of scoring.

Briefly, the plans of scoring were as shown in Table II.

TABLE II.

Summary of Plans of Scoring

Test	Score
Spelling	1 count for each correct answer and $\frac{1}{3}$ count for each blank. (nearest whole number)
Arithmetic	1 count for each correct answer.
Synonyms	1 count for each correct answer and 1 count deducted for each incorrect answer (blanks not counted).
Memory for Digits	1 count for each number entirely correct.
Proverbs	1 count for each correct answer.
Disarranged Sen ences	1 count for each correct underlining with 1 count deducted for each incorrect underlining.
Relation	1 count for each correct answer
Geometric	1 count for each figure 1 correctly placed, provided no other figure 1 appeared in the same design, and similarly, 1 count for each figure 2 correctly placed.
Following Directions	1 count for each direction correctly followed.
Narrative Completion	1 count for each blank satisfactorily filled.

The scores for each individual in each test will not be given as obtained by the above plan but instead they will be given in an altered form explained below. The scores are given in Appendix II.

V. The Reliability of the Scale

The reliability of a test may be expressed in two ways, either (1) by giving the probable error of a score in the units of the score, the probable error being the value of that error which is exceeded in amount by half the errors, or(2) in terms of the coefficient of correlation between two tests of the same kind. The probable error of a score as a measure of the reliability of a scale is comparable with other values of the probable error found in connection with the testing of other groups of individuals but it is not comparable with the probable error of the scores of other tests unless the units in the two scales measure the same increments of ability. This would not happen often and only accidentally. The reliability co fficient, as has been explained more fully elsewhere (see Ref. 11), ederived from measures of one group is not comparable with a reliability coəfficient for the same test derived from measures of another group unless the heterogeneity of the two groups is the same or nearly the same. In many instances, of course, this is not the case. The reliability coefficient of one test, however, is comparable with that

of another test when both are derived from measurements of the *same group*. The reliability coefficients are necessary under these conditions to show the relative reliabilities of two tests. They compensate the measures of reliability for inequalities of scale units.

We have therefore found both the probable errors and the reliability coefficients of each of the ten tests. The probable errors of scores in the several tests were found according to the method described at length in Ref. 11. This method is expressed by the formula,

$$\text{P. E.} = \frac{\text{Med. Dif}}{\sqrt{2}}$$

in which Med. Dif. is the median difference between scores by the same individuals in the two tests of the same kind. (A test of the first scale is called Test I; the corresponding test of the second scale Test II.) Before making the subtractions, however, it is necessary to have the scores of both tests in terms of either one or the other of the two tests, since these are quite often somewhat different; due to slight differences in difficulty, to practice effect, etc. For the purpose of evaluating the scores of one test in terms of the other, plots were made in which the scores in Test I were represented as abscissae (horizontally) and those of Test II as ordinates. The manner in which the scores in the two tests corresponded was then found by drawing in each plot a line of relation. This is such a line that the abscissa and ordinate of any point on it represent corresponding scores in the two tests. By inspection of the plots, it was deemed valid to draw a straight line of relation in all cases except that of the Narrative Completion Test, in which it was apparent that the true line of relation was markedly curved. In that case the curve of relation was drawn by the method we have called the method of correspondence by rank (see Ref. 7). In all cases except that of the Narrative Completion Test, the line of relation was obtained by finding the means, M_x and M_y, of the values of x and y and the average deviation, A. $D._x$ and A. $D._y$., of the distributions of values of x and y, and then drawing a line through the point (M_x, M_y) having a slope such that the tangent of the angle formed with the X axis = $\frac{\text{A. D.}_y}{\text{A. D.}_x}$

To find the score in terms of Test II which corresponds to any given score in terms of Test I, it is necessary merely to find the point on the relation line corresponding to the score in Test I and to note the

score in Test II at the left which corresponds to this point. The differences between the scores, in terms of Test II, are measured by the distances of the points of the plot above or below the line; in terms of Test I, by the distances to the right or left. The values of *Med. Dif.* were obtained by the method. (See Ref. 7):

$$\text{Med. Dif.} = .8453 \text{x} \text{Avg. Dif.}$$

That is,

$$\text{P. E.} = \frac{.8453 \text{ (Avg. Dif.)}}{1.414}$$

The values of the probable errors of each of the several tests were obtained first in terms of Test II and the corresponding values in terms of Test I were derived by dividing by the tangent of the angle of the line of relation. The values of the probable error in both terms are given in Table III.

TABLE III.

Reliability of the Tests

	Probable Errors		Reliability Coefficients	
	Scale I.	Scale II.	Single Tests	Double Tests
1. Spelling..................	1.49	1.45	.942	.970
2. Arithmetic...............	.74	.80	.871	.931
3. Synonyms and Antonyms..	1.96	1.86	.753	.8[illegible]
4. Memory for Digits........	1.04	1.21	.746	.855
5. Proverbs.................	1.17	1.02	.761	.864
6. Disarranged Sentences.....	1.28	1.76	.737	.849
7. Relation.................	1.50	1.86	.729	.843
8. Geometric...............	1.22	1.15	.805	.892
9. Following Directions......	.75	.97	.820	.901
10. Narrative Completion.....	5.43	3.80	.840	.913
			Total	8.877

The formula used for finding the reliability coefficients was:

$$r = 1 - \tfrac{1}{2}\left(\frac{\text{A. D.}_{(\text{dif's})}}{\text{A. D.}_{(\text{scores})}}\right)^2$$

which is a variation of the difference formula:

$$r = 1 - \tfrac{1}{2}\left(\frac{\sigma_{(y-x_y)}}{\sigma_y}\right)$$

This latter formula is the equivalent of the Pearson product-moment formula. (See Ref. 11.) In these formulae, $\text{A. D.}_{(\text{dif's})}$ and $\sigma_{(y-x_y)}$ are measures of the variability of the distribution of

differences between the scores of each of the 121 pupils in Test I and Test II, when the scores in Test I are evaluated in terms of Test II; and in which A. D.$_{(scores)}$ and σ_y are respectively corresponding measures of the variability of the distribution of *scores* in Test II.

The reliability coefficients thus found for each test are shown in the third column of Table III.

We were quite surprised to find the Spelling Test to be so much in the lead in this rating. However, the Spelling, Narrative Completion, and Synonym Tests had 50 elements while the other tests had only 25 or less. The Arithmetic, Following Directions, and Geometric Tests no doubt have an advantage over the others in that Test II was only slightly different from Test I.

The aim in duplicating the tests, as has been stated, was to make the second test in each case as nearly like the first as possible without actually copying it. This was done in order that the score in the second test would be as near as possible to a second score in the same test. It is possible that a second score in the same test would have been preferable for finding the reliability if it had been convenient to separate the two givings of the tests by a sufficient interval. Even this, however, would introduce new sources of error. Since the differences in difficulty between the two tests of a kind are not the same for all the pupils, the differences between the scores in the two tests tend to be greater than would be the case it the same test could be given twice, even without memory of the first testing, in which case the difference in the scores would be due merely to differences in disposition at the times of taking the first and second tests. For this reason, the values of the probable errors and reliability coefficients, considering only errors due to varying disposition, are really less than those given here.

Further consideration regarding reliability will be given later. These depend upon the values of inter-test correlations.

VI. Graduation of the Scale

Theoretical Considerations.—There are two aspects to the graduation of the scale. One deals with the proper combining of the scores of the several tests and the other with the finding of age norms, percentage norms, etc. The scores of each individual in the ten tests must first be combined into a single score, say a "point-score," and then those point-scores may be determined which are

normal for each of the given ages of childhood, or those which given percentages of adults may be expected to attain, etc.

In order that the scores of an individual in the several tests may be properly averaged, it is necessary to take account of the differences in value of the units of the scales of the several tests. If an increment of one problem in an Arithmetic score is in reality equal to an increment of four words in the score of the Synonym Test, to average the scores in the two tests just as they stand would be to give the Synonym Test four times as much weight as the Arithmetic Test. If, therefore it is desired to give equal weight to each test, the score of an individual in each test must be transmuted into other terms, say "points," such that *equal increments of ability* in each test receive *equal increments of points*. It is convenient, also, while assigning point values to the scores in the several tests, to arrange that correspondihg *amounts* of ability in the several tests shall receive corresponding *numbers* of points. The first of these conditions is essential and the second convenient for the purpose of averaging scores properly; both conditions are essential for the purpose of comparing scores in the several tests with one another. If it seemed reasonable to assume that for the individuals of a given group, the ability possessed by the upper 25% in any test was as much above that possessed by the upper 50% as that ability was above the ability possessed by the upper 75%, and if the ability in any one test was considered equal to that in any other test which was possessed by the same percentage of individuals, then the first of the above mentioned conditions would be complied with by representing the difference between upper 25% and upper 50% ability in each of the several tests by some number of points (say 10), and the difference between upper 50% and upper 75% ability in each of the several tests by that *same* number of points (10). And the second condition would be complied with by representing 50% ability in all of the ten tests by the same number of points (say 50) in which case, of course, 25% ability would be represented in each case by 60 points and 75% ability by 40 points.

We have been speaking of the equality of increments of ability, but such equality is a very indefinite thing. Equal increments of ability must be such as are measured by the same number of units of some kind. We have not been willing to grant that the steps of any test scale necessarily measured equal increments of ability.

Nor would we admit that any year's growth in ability is equal to every other year's growth. The growth of ability is supposed to retard eventually with age. In what units then will we say ability may be measured so that equal numbers of units measure equal increments of ability? In a previous article (Ref. 10) we have suggested that absolute units of ability be so defined that the distribution of abilities of all adults *will be normal* (in the technical sense). This would mean that those percentages of adults which were considered as possessing abilities which marked successive steps on an absolute scale of ability were the same percentages as those of the normal probability surface which corresponded to successive units of the base. Until such time, however, as a very large number of unselected adults have been tested, such a criterion of equality of units of ability will be unavailable. In lieu of such a criterion, an alternative method was used.

The Procedure Used for Determining Equality of Increments of Ability.—Although we have felt that the units in one part of a single test scale were very apt to be of greater value than those in some other part, it is quite probable that if the upper units of some test scale must be considered as measuring greater increments of ability than the lower units, the opposite probably might be considered true of some other test scale, so that taking the test scales all together, the median value of the units in one part may be considered as equal to the median value of the units in any other part. Proceeding upon that hypothesis, the most probable true form of the distribution of abilities of the 121 pupils was determined by obtaining a composite of the separate distributions for the ten tests as follows:

1. The score attained in each test by the 30th individual in rank (beginning with the lowest) was assigned a preliminary point-value of 40 points and the score attained by the 90th individual in rank was assigned a preliminary point-value of 60 points.*

2. Tentative point values corresponding to all the other scores were then determined in such a manner that the units in all parts of the test scale were represented by equal increments of points. This was accomplished graphically in each case by drawing a straight line.

3. From the smooth curves of distribution of test scores were

*These scores were not the *actual* scores of those individuals but the scores corresponding to them on smooth curves through the distributions of consecutive corses.

then determined the scores attained by the 3rd, 9th, 15th, 60th, 105, 111th, and 117th individuals in rank order. These points in the distribution curves were believed to best reveal any skewness of the distribution.

4. The preliminary point values corresponding to the scores attained by the 3rd individual in each test distribution were then ascertained. These were then plotted in order of magnitude and a median value determined by means of a smooth curve through the plotted points. This median point value was 24.4. The other median point values were as follows:

Individual in order:	3,	9,	15,	(30)	60,	(90)	105,	111,	117
Point value	24.4	29.7	33.3	(40)	50.1	(60)	66.7	70.1	75

It should be stated that these values indicate that the distribution of abilities of the 121 pupils approximately normal.

5. Since the median of the preliminary point values obtained by the 3rd individual in rank in the several test distribtuions was 24.4, this value may be assumed to be the most probable true value, in terms of our established absolute units, of the ability in any test which the 3rd individual in rank order attained. The score in each test attained by the 3rd individual in rank order (by the curve) was then given, therefore, the corrected point value 24.4. Similarly the score in each test attained by the 9th individual was then given the corrected point-value 29.7, etc.

6. In order to determine the corrected point value to be similarly assigned to all the other scores in each test, a graph was made for each test in which the preliminary point values corresponding to the scores attained by the 3rd, 9th, etc., individuals were plotted as ordinates and the new point values, 24.4, 29.7, etc., plotted as abscissae. A smooth curve was then drawn through the series of plotted points. This curve was then taken as showing the relation between the preliminary and corrected point values corresponding to each score in the test From this curve for each test were taken the corrected point values corresponding to each score. These are shown in Table IV. They no doubt represent the nearest approach that can be made to a true absolute point scale.

Considerations with Regard to Weighting and Combining the Scores.—After finding the corrected point values corresponding to each test score, the scores of each pupil in each test were transmuted into terms of points and the total score found for each. These are given in Appendix II.

This method of combining the scores resulted in equal weight being given to each test. No doubt some of the tests are more significant than others in the measurement of general ability, however we conceive it. Unreliability of a test, of course, lowers its significance. Other aspects of significance depend upon the conception of general ability. If a test is considered as measuring general ability only to the extent to which the factors entering into the ability tested are common to other abilities, both as to number of factors and as to number of abilities to which they are common, then the degree to which a test may be considered as measuring

TABLE IV

Showing the Number of Points Corresponding to Each Score in Each Test

Score	Spelling Points	Arith. metic Points	Synon- yms Points	Digits Points	Proverbs Points	Disar. Sentences Points	Relation Points	Geomet. Points	Fol. Direc. Points	Narra. Comple- tion Points
0		21		20	32	23	25	21	23	26
1		24		20	34	28	26	22	27	26
2		27		21	35	33	27	24	31	27
3		30		22	36	38	28	25	35	27
4		33		23	37	43	29	27	39	28
5		36		24	39	47	31	28	43	28
6		39		26	40	50	32	29	48	29
7		42		28	41	53	33	31	52	29
8		45		31	43	56	35	32	56	30
9	20	48		35	44	58	37	34	60	30
10	21	51		38	45	61	40	36	64	31
11	22	54		42	47	63	42	38	68	31
12	23	56		45	48	66	45	41	72	32
13	24	59		48	49	68	47	43	76	32
14	25	62		51	50	71	50	46	80	33
15	26	65		55	52	73	52	49		33
16	26	68		58	53		55	52		34
17	27			61	55		57	56		34
18	28		20	64	56		60	59		35
19	29		22	68	57		62	63		35
20	30		23	71	59		65	66		36
21	31		25				67	70		36
22	32		26				70	74		37
23	32		28				72			38
24	33		29				75			38
25	34		31				77			39
26	35		32							39
27	36		34							40
28	37		35							41
29	38		37							41
30	38		38							42
31	39		40							43
32	40		41							44
33	41		43							45
34	41		44							46
35	42		46							47
36	43		47							48
37	44		49							49
38	45		50							50
39	46		51							51
40	47		53							52
41	48		54							53
42	49		56							55
43	50		57							56
44	51		59							57
45	52		60							58
46	53		62							59
47	55		63							61
48	56		65							62
49	58		66							63
50	60		68							64

general ability is expressed by the amount of "correlational spread" of the test, to use McCall's expression, by which is meant the sum of the intercorrelations of the test with other tests comprising a fairly representative collection, each presumed to involve factors common to the others. The last qualification is necessary since, if the group of tests is too restricted in kind, certain 'specific' abilities may be common to too large a proportion of the tests and thus vitiate the criterion of general ability.*

On the other hand if a test is considered as contributing to the measure of general ability if it measures an ability that may be considered valuable in aiding the individual to adjust himself to the new problems and conditions of life, whether such ability has few or many factors in common with others; then it is not proper to use only the criterion of correlational spread. Two possible alternatives suggest themselves. If there were available for the individuals tested a satisfactory criterion of their powers of adaptation to the new conditions and problems of life, in the nature of a measure of economic or scholastic success, then it would be necessary merely to weight the tests according to the regression equation method, so as to obtain the best correlation of the composite score with the criterion. In lieu of such a criterion, the tests might be weighted according to a combination of the weights assigned by a number of judges. In this study, for instance, the results of all the tests except that of Memory for Digits correlated uniformly highly with each other. The Digit Test, which showed a reliability not the least among the ten tests, stood quite apart from the other tests in showing low correlations with all of them. According to the criterion of correlational spread, this test would be weighted very much lower than any of the others. According to either of the criteria pertaining to the second conception of general ability, however, the Digit Test might perhaps deserve a weight more nearly the amount of the others.

*Some mathematical reasoning bearing on this point is given in Appendix III, 1 and 2.

McCall used this criterion in his study (Ref. 5). Another criterion which he also used was the correlation of each test with "Composite," a measure obtained by combining the scores of all the tests (with some exceptions) after weighting each according to *a priori* considerations as to the value of the tests. Although the correlations of the several tests with Composite appear to have been determined by McCall by separate calculations, it would have been possible to obtain the values of these correlations with Composite more simply from the values of the inter-test correlations. The necessary procedure is given in Appendix III, 3.

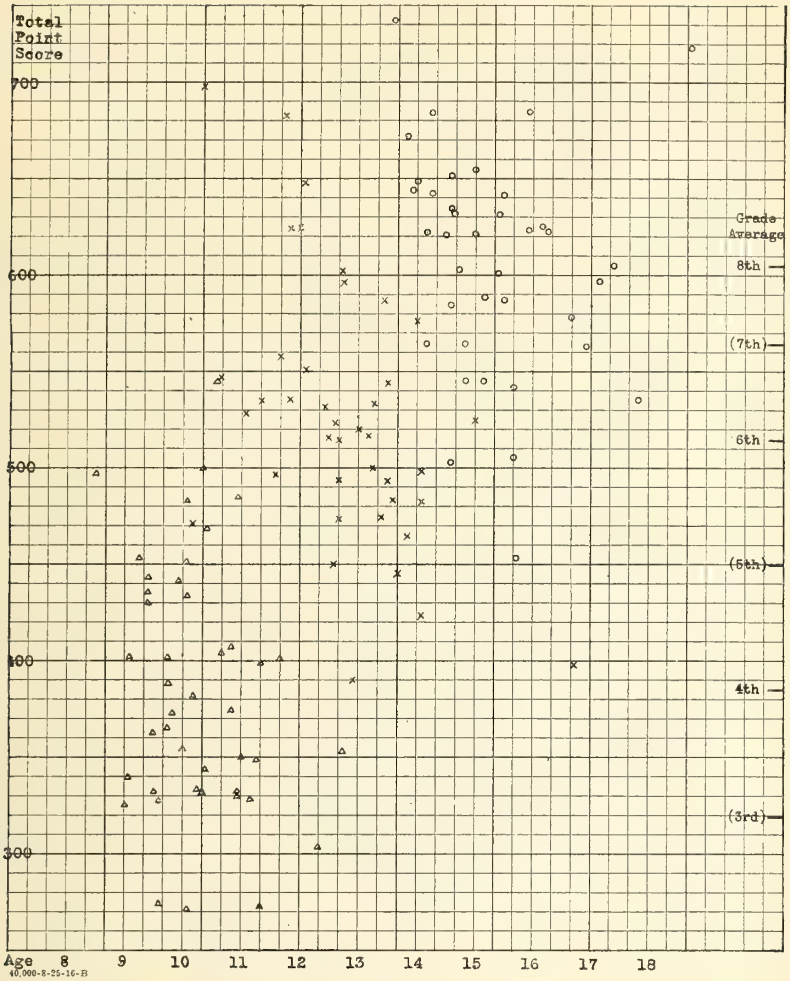


FIG. 1

Showing the Relation between Total Point Score and Age

In this study we are inclined more to the second conception of general ability mentioned. It was not feasible in this study, however, to use either of the criteria appropriate to this conception. To weight the tests according to reliability alone, it would be necessary to weight each inversely in proportion to the square of the probable error (the probable errors being in comparable terms). (See Merriman, Ref. 6, p. 95.) Such procedure, however, prac-

tically implies that all the tests aim to measure the same thing. But since they do not, any weighting given to compensate for different degrees of reliability, necessarily also emphasizes the effect of certain particular abilities and is to that extent undesirable.

For these reasons we have combined the test scores without weighting them.

Finding Age Norms in Terms of Point Scores.—For finding age norms, a plot was made. (See Fig. 1.) One point pertains to each pupil. The abscissa of each point represents the pupil's age and the ordinate his total point score. In order to find the score which would be considered normal for 10-year-olds, the average score was found of all pupils of ages from 9 years, no months, to and including 11 years, no months; for 11-year-olds, the average score was found of all pupils of ages 10 years to and including 12 years, etc. The norms thus found were as shown in Table V. These values were then plotted. (See Fig. 2.) To our surprise, the points representing the norms for ages 10 to 14 lay in almost a perfectly straight line, which suggests that they are fairly reliable, at least, for the school population tested. This was not expected considering the gaps left by omitting the fifth and seventh grades from the group tested. The norms for years 15, 16, 17 may be seen to fall below the line, the latter two quite markedly. This was to be expected, of course, since the pupils of these ages were selected, being retarded in their schooling. While the true norms for these ages are doubtless above the average values obtained, it was not deemed proper to continue the straight line. The line was therefore curved off to the right as shown. We must regard the norms for the ages above 15, as being only roughly approximate.

TABLE V.

Showing Age Norms in Point Scores

		8	9	10	11	12	13	14	15	16	17	18	19
Age:		8	9	10	11	12	13	14	15	16	17	18	19
Point: Score:	Observed:			404	446	487	527	566	583	550	584		
Norms:	Smoothed:	324	364	405	445	486	526	566	600	624	638	647	650

Completing the Absolute Point Scale.—We have previously (see Ref. 10) given the name, Coefficient of Brightness, to the quotient that would be obtained by dividing the measure of the absolute amount of mental ability of any individual by the measure of the absolute amount of mental ability which was normal for the age of that individual. This means, of course, that the measures of

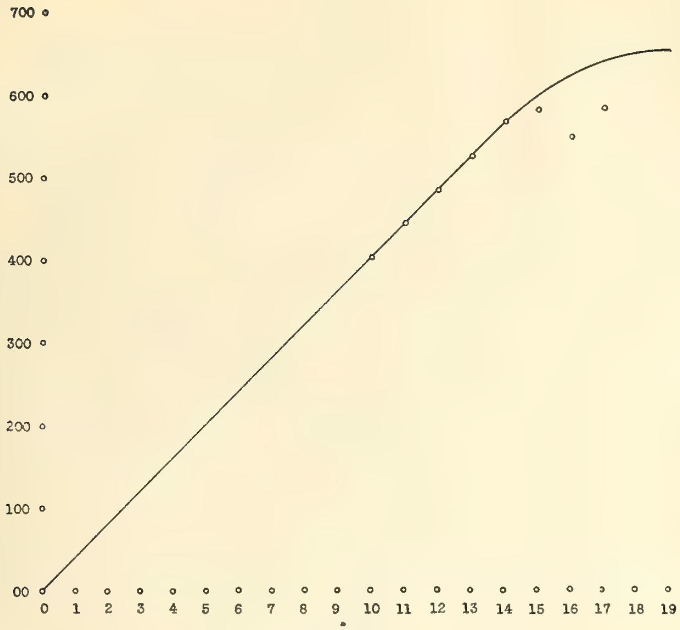


FIG. 2
Showing a Smooth Curve through the Age Norms of Total Point Scores

mental ability must be in such terms that not only will equal increments of ability be measured by equal increments of the scale, but twice as many units on the scale will represent twice as much ability, etc. In other words, zero of the scale must represent just absence of ability. Before it was possible for us to find the coefficients of brightness of the pupils tested in this case, therefore, it was required to note what correction was necessary in the scale of points in order that the number of points representing the ability of age 0 would be 0. The ages for which we may presume to have obtained fairly reliable norms are only those from 10 to 14. Inasmuch, however as the increments of points between the norms for these ages are almost exactly the same, it was regarded as proper to assume for present purposes that if continued, the line through the norms would be straight the rest of the way to age zero. It was then necessary to note what number of points thus corresponded to age zero, this number to be considered the true absolute zero of the final point scale. To our further surprise, it was discovered that by calling the yearly increment of points (below 14) approx-

imately 40.5 a line, which would pass as nearly as any other through the five norms, actually reached zero age at zero of the point scale. This, of course, was an entirely accidental coincidence and not at all necessary. It merely saved us the obligation of subtracting or adding a constant to each of the corrected point values assigned to the several test scores in order to obtain the final point values constituting the completed Absolute Point Scale.

The Determination of the Coefficients of Brightness.—Since the point values in which the scores of the pupils were expressed proved to be those of the Absolute Point Scale, in order to find the coefficients of brightness of each pupil, it was necessary merely to divide the total point score of each by the score which was normal for his age. The norms for the fractional ages were taken from the curve in Fig. 2. The coefficients of brightness thus found are given in Appendix II.

Appendix I.

Sample Extracts of Tests:

Test 1: Spelling

1. forenoon	fournoon	(F)
2. intrest	interest	(S)
3. neighber	neighbor	( )
4. concider	consider	( )
5. entertain	entertane	( )
etc.	etc.	etc.

Test 2: Arithmetic

1. If a boy has 10 cents and then earned 5 cents, how much did he have then? . . . . ( ) cents

7. How many years will it take a glacier to move 1000 feet at the rate of 100 feet a year? . . . . ( ) years

15. A ship has provision to last her crew of 50 men 6 months. How long would it last 30 men? . . . . ( ) months

Test 3: Synonyms and Antonyms

1. large	big	(S)
2. decrease	increase	(O)
3. empty	vacant	( )
4. knowledge	ignorance	( )
50. conservative	radical	( )

Test 4: Memory for Digits

1. 4739	( ) ( ) ( ) ( )
2. 2854	( ) ( ) ( ) ( )
3. 7261	( ) ( ) ( ) ( )
4. 31759	( ) ( ) ( ) ( ) ( )
5. 42385	( ) ( ) ( ) ( ) ( )
6. 98157	( ) ( ) ( ) ( ) ( )

VII. Overlapping of Ability Between Grades

The points in Fig. 1 belonging to pupils in the eighth grade were made as circles, those belonging to pupils in the sixth grade, crosses, and those belonging to pupils in the fourth grade, triangles. It will be noted that there is considerable overlapping between the grades even though they are not consecutive.

The average score of the fourth graders is 385, of the sixth graders, 514; and of the eighth graders, 605. Suppose we call the norm for the third grade 320, the norm for the fifth grade 450, and for the seventh grade, 565, as shown in Fig. 1. We then find 8 fourth graders out of 43 above the fifth grade norm. Presumably these could do satisfactory fifth grade work. We find 1 of these 8 above the sixth grade norm. And we find 4 fourth graders below the third grade norm. The scattering of the three grades is shown in Table VI.

TABLE VI

Showing the Overlapping between the Grades

Norms:		3rd	4th	5th	6th	7th	8th
Fourth Grade (43)	4	19	12	7	1		
Sixth Grade (40)			5	11	15	4	5
Eighth Grade (38)				3	5	10	20

Another rather interesting fact concerning the distributions of scores, particularly in the sixth grade, is that there is a tendency for the more mature* pupils, intellectually, to be the younger ones. It would seem from this and the many other similar investigations that this is invariably the case. The most mature pupil in the sixth grade is, in fact, next to the youngest, while the oldest is next

*It has been necessary in this case to avoid the use of the ambiguous word, intelligence, which is used by nearly all writers on mental testing to mean both maturity, irrespective of age, and brightness—maturity with respect to age. The statement that, in a single grade, the youngest are also the most intelligent, according to the second meaning, would be a mere platitude. This would be true even if there were zero correlation between age and maturity.

to the least mature. In the eighth grade, also, the youngest is more mature, intellectually. There is, in other words, a negative correlation between age and maturity in the single grades. If pupils were graded according to intellectual maturity only, there would be no appreciable correlation, positive or negative, between age and maturity in a single grade. The fact of negative correlation, therefore, suggests strongly that some bright pupils (mature but young) have been held back by the inelastic system of grading and that dull pupils have been promoted beyond their ability. This is one of the evils which mental testing should eventually remedy.

VIII. The Refinement of the Scale

Finding the Order of Difficulty of the Elements of the Tests. While not essential, it is nevertheless very desirable to have the elements of each test arranged in the order of difficulty. The relative degrees of difficulty of the elements of a test are, of course, probably not the same for any two individuals. The best arrangement, however, is probably the order of the elements according to the number of individuals who pass each, beginning, of course, with the easiest. In order to determine this ranking, the number of individuals who failed in each element was found during the scoring, for the Spelling, Arithmetic, Synonyms, Proverbs, and Relation Tests. To give an idea of the distribution of difficulties of the elements of these five tests in Scale 1, Fig. 3 was made. The horizontal position of each circle represents the number of individuals who failed in a given element. The circles at the left, therefore, represent easy elements. It is apparent from this as well as other sources that the Spelling Test is too easy for this group of individuals. The elements should be of such difficulty that the median element, in difficulty, is passed by about 50% of the group. The Synonym Test is somewhat too easy. The Arithmetic problems appear to fall into two distinct groups in difficulty. Problems of medium difficulty should be substituted for some of the others. The distributions of difficulty in the Relation and Proverb Tests are, perhaps, fairly satisfactory.

The Diagnostic Value of the Single Test Elements. It is not deemed within the scope of this study to investigate the value of each element of each test, as for example, a single problem in Arithmetic, as a measure of general ability such as is measured by the total point score of an individual, or of general arithmetical ability as measured

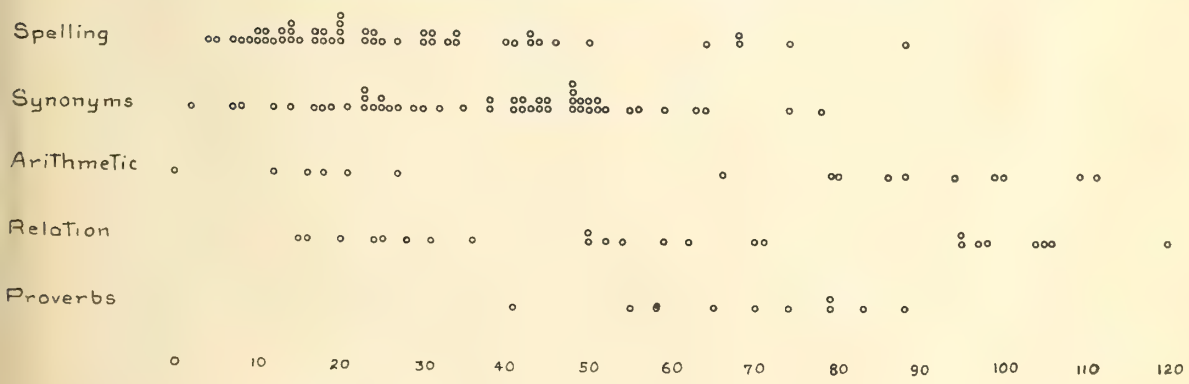
by the arithmetic score. However, as suggestive of means by which this may be done, we have examined the sixteen elements of Arithmetic Test I with the view to discovering which were the most suitable to be included in a test designed to be part of a scale for measuring general ability.

The method employed was as follows: The 121 individuals were first ranked in order of their total point scores. The papers of the 121 individuals in the Arithmetic Test I were then arranged in the same rank order. The sequence of passings and failings of Problem 1, Problem 2, etc., were noted. These are represented in Fig. 4 for each of the 16 problems and for the 121 individuals. There is one dotted line for each problem; each dotted line contains 121 units, one for each individual in the order of their total point scores. The presence of a unit of line indicates a problem correct; the absence of a unit line indicates a problem failed. The lines are arranged in the order of difficulty of the problems beginning with the easiest, the numbers of the problems represented are given at the left. The number of passes for each problem is represented by the position of a small circle on the line. In this figure the relative values of the problems as measures of general ability are shown by the relative amounts of overlapping of passes and failures—the greater the overlapping, the less the diagnostic value of the problem.

If the range of abilities of the individuals tested had been sufficiently broad so that the complete range of overlapping was represented for each problem, it would be a comparatively simple matter to express the relative diagnostic value of each problem by a single

FIG. 3

Showing the Distributions of Difficulties of the Elements of Five Tests



number. For example, let us suppose that no individuals having measures of general ability above those represented in the figure would fail in either of Problems 7 and 8 and that no individuals having measures of general ability below those represented would solve either of these problems. If there were no overlapping in either case a full line would extend just to the circle representing the total number of passes, these being 94 and 55 respectively. Since in line 7 there are 10 failures before and 10 passes after the circle, we could represent the amount of overlapping by the number 10. Similarly since in line 8 there are 23 failures before and 23 passes after the circle, could represent the amount of overlapping in this case by the number 23. A rank order of the problems according to these numbers would, under the conditions mentioned above, give a serviceable indication of the comparative diagnostic values of the problems.

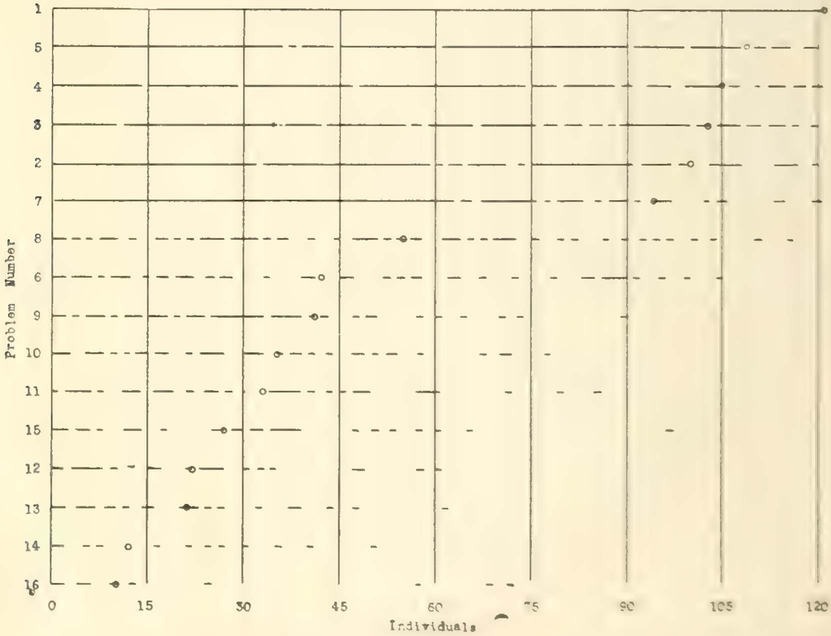


FIG. 4

Showing the Rank Order in Intelligence of the Individuals who Passed Each Problem of Arithmetic Test I.

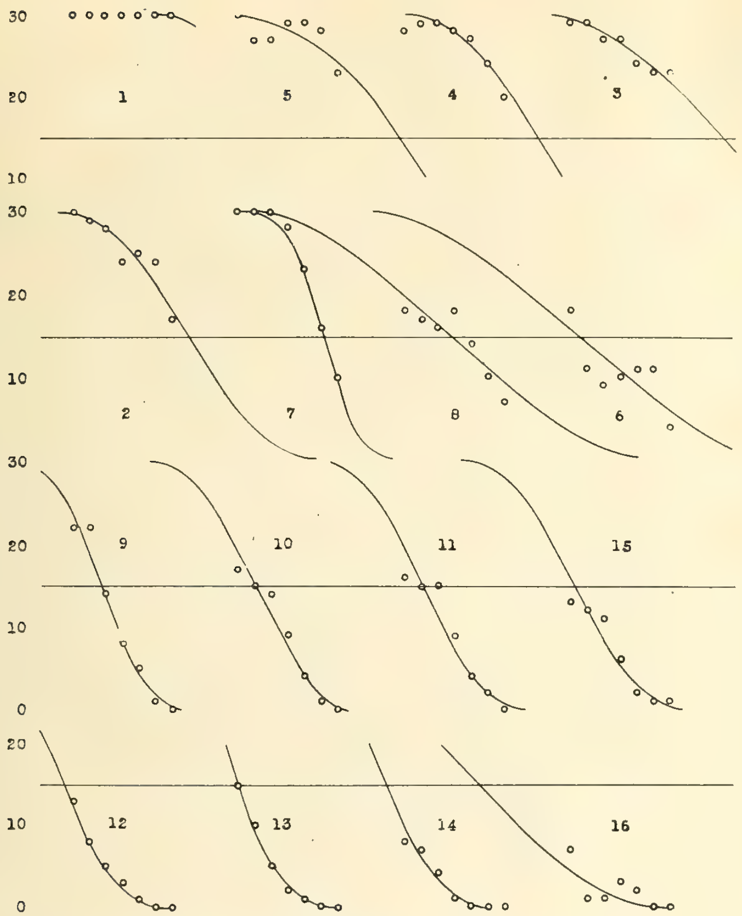


FIG. 5
Showing the Diagnostic Value of Each Problem of Arithmetic Test I

Inasmuch, however, as only a portion of the overlapping is represented in each case, it becomes necessary to adopt further means of ascertaining the relative diagnostic values. A method offering greater refinement is illustrated in Fig. 5. In this figure each group of seven circles pertains to one problem. The heighs of the first circle, according to the scale at the left, represents the number of passes by the first 30 individuals in rank order. The height of the second circle represents the number of passes by individuals 16 to 45, inclusive, the third circle, individuals 31 to 60, etc., embracing 30 individuals in each group, the last group being 91 to 120. There is, of course, a tendency in each case for the succeeding numbers of passes to decrease. Theoretically, there should be a tendency for the circles to lie in a smooth curve of the form of an ogive. The steepness of the curve would indicate the degree of diagnostic value of the problem. The merit of this method is that it is possible in many cases to obtain a fairly good idea of the true slope of the curve for a given problem from only partial data. Curves have been drawn in what was judged by the eye to be approximately the true position of the curve. The problems may now be ranked in diagnostic value according to the slope of the curve at the 50% point. Thus it will be seen that the diagnostic value of Problem 8 appears to be the best. The value of Problem 1, of course, cannot be found. Quite possibly it would be as good as the average for lesser degrees of ability. For this group of individuals, of course, it has no value except perhaps for illustrative purposes. With the exception of Problems 6 and 7, possibly of 3 and 16, it would seem that the diagnostic values of the problems may be considered satisfactory.

It should be noted that the horizontal position of the point at which the curve crosses the 50% line affords a refined measure of the degree of general ability to which the ability to solve the problem corresponds. Thus, Problem 8, for example, may be considered "standard" for the degree of general ability slightly less than that of the 90th individual in rank order, say of the 93rd; or of 600 points. This method is practically that suggested for the standardization of tests of the Binet Scale (see Ref. 15). To thus express the degrees of difficulty of the several problems in terms of the Absolute Scale would assist in making the increments of difficulty between the problems equal. Such an amount of refinement however, is not considered to be of great value until more of the obvious defects of the scale have been eliminated.

To determine the relative values of the problems as measures of "arithmetical ability," it would be necessary, of course, merely to rank the papers in the order of the arithmetic scores instead of the total scores.

IX Inter–Test Correlations

Considering the fact that double measures were obtained of each ability tested and that the number of individuals was comparatively large, it was deemed valuable to obtain the inter-test correlations.

These are given in Appendix IV. They are correlations between measures obtained in each case by combining the results of the two tests of a kind. There are given both the raw coefficients and those corrected for attenuation due to errors of measurement. The formula used for correcting for attentuation was as follows (see Ref. 3).

$$r_{ab(corrected)} = \frac{r_{ab(raw)}}{\sqrt{r_{aa}\, r_{bb}}}$$

It is considered necessary to leave the discussion of the intercorrelations to a later article.

X. Further Considerations Regarding Reliability

The Reliability Coefficient of the Point Scale. To find the reliability coefficient of correlation between Scale I and Scale II, we may proceed as follows. Let us call the ten tests of Scale I a_1, b_1, c_1, - - - j_1, and those of Scale II, a_2, b_2, c_2, - - - j_2. Then by the formula (See Ref. 13) for the correlation of the sums of several variables, the standard deviations of the distributions of scores in the several tests having been made equal,

$$r_{(a_1+b_1+c_1+ \cdots +j_1)\,(a_2+b_2+c_2+ \cdots +j_2)} =$$

$$\frac{r_{a_1a_2}+r_{a_1b_2}+r_{a_1c_2}+ \cdots +r_{b_1a_2}+r_{b_1b_2}+r_{b_1c_2}+ \cdots +r_{c_1a_2}+r_{c_1b_2}+r_{c_1c_2}+ \cdots}{\sqrt{10+2(r_{a_1b_1}+r_{a_1c_1}+\cdots+r_{b_1c_1}+\cdots)}\ \sqrt{10+2(r_{a_2b_2}+r_{a_2c_2}+\cdots+r_{b_2c_2}+\cdots)}} \quad (1)$$

$$= \frac{\text{(Sum of 10 reliability coefficients, } r_{a_2a_2}\text{, etc.)} + 2\text{(Sum of 45 coef. of intercorrelation, } r_{a_1b_2}\text{, etc.)}}{\sqrt{10+2\text{(Sum of 45 coefs. of intercor., } r_{a_1b_1}\text{, etc.)}}\ \sqrt{10+2\text{(Sum of 45 coefs. of intercor., } r_{a_2b_2}\text{, etc.)}}}$$

But since the correlations, $r_{a_1b_1}$, $r_{a_1b_2}$, and $r_{a_2b_2}$, tend to be equal, we may take the sums of each of the three sets of 45 coefficients of

correlation to be equal to one another. We may therefore simplify the equation thus:

$$r_{(a_1+b_1+c_1+\ldots+j_1)\ (a_2+b_2+c_2+\ldots+j_2)} = \frac{\Sigma r_{xx}+2\Sigma r_{xy}}{10+2\Sigma r_{xy}} \qquad (2)$$

in which $r_{xy}=r_{aa}+r_{bb}+r_{cc}+\ldots$

and $\quad r_{xx}=r_{ab}+r_{ac}+r_{ad}+\ldots+r_{bc}+\ldots$

Since the intercorrelations were found between *double* measures in each test, it is necessary to express the reliability coefficients also in terms of double measures. As these were found in terms of single measures, each has been transmuted into terms of double measures by means of the formula,*

$$r_{2a2a}=\frac{2r_{aa}}{1+r_{aa}} \qquad (3)$$

in which r_{2a2a} and r_{aa} are respectively the correlations between double measures and between single measures of any abilities. The reliability coefficients in each test in terms of double measures are shown in the fourth column of Table III. Their sum (Σr_{xx}) is shown to be 8.877. The sum of the 45 coefficients of intercorrelation multiplied by 2 = 56.478 ($=2\Sigma r_{xy}$). Solving formula 2 above,

$$r_{(\text{Scale I}_2)\ (\text{Scale II}_2)}\ \frac{8.877+56.478}{10+56.478}=.983.$$

This is when Scale I_2 and Scale II_2 are considered as double scales. To find the reliability coefficients of correlation between Scale I and Scale II as single scales, *i. e.*, each composed of the tests truly comprising it (call these Scale I* and Scale II*), then, according to the formula,†

$$r_{aa} = \frac{r_{2a2a}}{2-r_{2a2a}}$$

in which r_{aa} and r_{2a2a} have the same meanings respectively as before,

$$r_{(\text{Scale I}_1)\ (\text{Scale II}_1)} = \frac{.983}{2-.983} = .967$$

This, then, is the reliability coefficient of the Point Scale.

The Probable Error of the Point Scale

As has been shown (see Ref. 11)

$$r=1-\frac{\sigma_e^2}{\sigma_{dist}^2} \qquad (1)$$

*This formula is a corrollary to formula 1 above.

†The inverse of formula 3.

in which r is the reliability coefficient of correlation between two series made up of pairs of measures, a_1, a_2; b_1, b_2; c_1, c_2; etc.; in which σ_e is the standard deviation of the errors of measurement of a, b, c, etc.; and in which $\sigma_{dist.}$ is the standard deviation of the distribution of values, a, b, c, etc., in either series.

From equation 1 it follows that

$$\frac{\sigma_e^2}{\sigma_{dist.}^2} = 1 - r \text{ whence } \sigma_e^2 = (1 - r)\,\sigma_{dist.}^2$$

$$\sigma_e = \sqrt{1 - r}\,\sigma_{dist.} \qquad \text{and} \qquad \text{P. E.} = .6745\sqrt{1 - r}\,\sigma_{dist.} \tag{2}$$

in which *P. E.* is the probable or median error of measurement.

If, now, we consider r to be the reliability coefficient of the Point Scale, P. E. as the probable error of measurement by either Scale I or Scale II, and $\sigma_{dist.}$ as the standard deviation of the point scores by the same scale, then we may solve equation 2 for the probable error of the scale. The standard deviation of the distribution of scores by the scale (average of both) was found to be 111 points. Solving equation 2,

$$\text{P. E.} = .6745\sqrt{1 - .968} \times 111 = 13.7$$

The Probable Error of the Point Scale, therefore, is 13.7 points.

This is 2.7% of the median score (500) of the whole group, or 3.0% of the total range of scores (461).

To view the reliability of the scale from another angle we may determine as nearly as possible the probable error of a mental age by the scale. Thus, if we may assume that the reliability co-efficient of correlation between mental ages by the scale is approximately equal to the reliability coefficient of correlation between the point scores and that the distribution of mental ages by the scale is approximately equal to the distribution of ages, then we may let P. E., in formula 2 above, represent the probable error of a mental age by the scale, r represent the reliability coefficient of correlation between mental ages by the two scales, and $\sigma_{dist.}$ represent the standard deviation of the distribution of mental ages. Then substituting in equation 2 the approximate values of these quantities, the standard deviation of the distribution of ages being 28.1 months, we have

$$\text{P. E.} = .6745\sqrt{1 - .967} \times 28.1 = 3.44$$

The Probable Error of a mental age by the Point Scale may be considered, therefore, as approximately 3½ months. This is practically the same as the probable error of a mental age by the Stanford Revision of the Binet Scale (see Ref. 11).

XI. Comparisons with School Mark and Amount of Schooling

Correlation of Total Score with School Mark. The teacher of each of the grades furnished for each pupil a final mark representing the relative character of his school work for the year. As the marks given by the different teachers were not comparable, a separate correlation was made for each grade between the school mark and the total point score. The coefficients were found for the fourth, sixth, and eighth grades to be respectively .80, .41, and .50. Since we have no measures of the reliability of the teachers' marks we are unable to determine the probable true correlation between intelligence, as measured by the scale, and school performance.

The Relation of the Coefficient of Brightness to the Amount of Schooling. The pupils were asked to tell the length of time they had spent in each grade. From these data, the total amounts of time spent in school was found for each pupil. For convenience, the 121 pupils were then classed together in groups according to the amount of retardation or advance.* The number of the pupils of each class are shown in Table VII. Here again we have no measure of the reliability of the reports of schooling and are therefore unable to determine the value of the results. There is, however, a very definite tendency for advanced pupils to obtain high coefficients of brightness and for retarded pupils to obtain low coefficients of brightness.

TABLE VII

Showing the Amounts of Retardation and Advance and their Relation to the Coefficients of Brightness. (Data from 10 pupils were missing.)

	Retarded			At	Advanced	
	3 yrs.	2 yrs.	1 yr.	Grade	1 yr.	2 yrs.
Number of Pupils	2	8	33	56	8	4
Average Coef. Brightness	77	88	96	102	103	114

*These terms refer here merely to the taking of more or less than the normal time to reach the present grade; they are irrespective of the pupils' ages of entrance.

Test 5: Proverbs

Pr verbs

(3) Make hay while the sun shines.
(   ) In a calm sea every man is a pilot.

Statements to explain the proverbs.

1. Deeds show the man.
2. Leadership is easy when all goes well.
3. Make the best of your opportunities.

Test 6: Disarranged Sentences

1. name a John is boy's......................................(true   false)
2. sun morning the the in sets....................................(true   false)
3. trees birds nests the in build.................................(true   false)

Test 7: Relations

1. hand : arm : : foot : (   )
2. hat : head : : thimble : (   )
23. education : ignorance : : (   ) : poverty

1. 1 leg, 2 toe, 3 finger, 4 wrist, 5 elbow.
2. 1 finger, 2 needle, 3 thread, 4 hand, 5 sewing.
23. 1 laziness, 2 school, 3 wealth, 4 charity, 5 teacher.

Test 8: Geometric Test

(*Designs were presented composed of two or more geometrical figures—circles, rectangles, and triangles—overlapped.*)

1. Place a figure 1 so that it will be both in the rectangle and in the circle.
7. Place a figure 1 so that it will be in both circles, in the triangle, and in only one rectangle.

Test 9: Following Directions

(*A page of Woodworth and Well's Cancellation Test was supplied each pupil.*)

2. In line 1 [of the forms] place a figure 1 in the first star and a figure 2 in the second circle.

. In line 5, place a figure 7 in the form which follows the same kind of form as that which follows it.

Test 10: Narrative Completion

Once upon a .... there was a y........ .... who was very p........
He went from place to .... trying to find ....

Appendix II.

Showing the Point-Scores of each individual in each test. (Pupils 1 to 41, eighth grade; 51 to 96, fourth grade; and 101 to 146, sixth grade.)

Pupil	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16
Spell.	77	55	68	56	63	64	59	62	65	57	54	70	72	53	63
Arith.	76	54	67	49	51	76	56	58	73	56	72	60	64	44	62
Synon.	76	58	64	50	57	67	72	59	57	60	56	75	69	47	64
Digit.	80	76	41	45	43	62	53	50	41	50	62	32	45	30	36
Proverb.	81	54	65	35	46	62	65	54	68	60	52	68	71	35	62
D. Sen.	68	72	63	63	68	75	63	72	60	59	71	53	72	59	63
Rel'n.	61	53	71	44	53	49	69	69	67	69	67	75	73	41	63
Geom.	52	55	58	55	58	58	66	69	74	66	64	58	61	45	61
Fol. D.	67	58	65	48	56	51	67	61	58	72	58	75	75	43	56
Compl.	80	62	64	58	68	58	65	68	68	72	48	68	83	56	58
Sums	718	597	626	503	563	622	635	622	631	621	604	634	685	453	588
C. B.	109	93	100	86	89	108	108	99	103	103	102	108	110	74	96

Pupil	17	18	19	20	22	23	24	25	26	27	28	29	30	32	33
Spell.	64	70	67	65	51	64	56	64	54	72	65	72	61	54	56
Arith.	66	58	49	64	58	64	70	56	66	64	70	56	64	72	51
Synon.	58	67	65	64	56	61	55	54	65	68	68	59	54	66	65
Digit.	48	62	50	60	45	43	30	57	76	57	41	66	30	64	57
Provb.	52	71	60	65	52	52	52	54	60	81	65	58	56	68	56
D. Sen.	67	63	60	70	68	64	53	73	65	63	68	66	71	75	58
Rel'n.	67	61	55	75	46	61	55	53	46	61	71	42	60	69	55
Geom.	55	69	52	64	64	40	49	66	66	61	64	61	64	55	72
Fol. D.	58	67	38	58	53	53	58	53	63	73	61	61	48	65	73
Compl.	54	63	50	56	43	63	64	55	44	72	69	60	57	66	66
Sums	589	651	546	641	536	565	542	585	605	672	642	601	565	655	623
C. B.	98	111	91	104	84	98	88	99	95	120	111	99	95	109	100

Pupil	34	35	36	37	38	39	40	41	51	52	53	55	56	58	59
Spell.	64	64	64	54	68	51	70	64	27	32	39	31	53	46	[illegible]
Arith.	58	60	67	77	70	67	70	79	44	27	39	34	19	46	[illegible]
Synon.	50	53	74	48	72	63	46	73	19	35	32	32	25	43	29
Digit.	55	45	75	75	53	36	64	78	41	36	68	50	45	50	36
Provb.	46	50	68	56	56	52	71	75	40	31	35	24	31	31	31
D. Sen.	49	59	68	70	68	65	65	63	48	38	40	25	48	49	42
Rel'n.	58	46	61	56	60	75	75	82	26	32	46	29	30	38	38
Geom.	29	52	72	66	61	58	66	72	29	36	41	43	25	45	[illegible]
Fol. D.	43	61	72	53	70	51	58	75	29	29	46	41	29	41	[illegible]
Compl.	54	56	62	66	66	60	64	71	27	36	44	35	35	52	38
Sums	506	546	683	621	644	578	649	732	330	332	430	344	340	441	[illegible]
C. B.	83	92	119	107	115	92	115	133	75	87	113	82	93	110	[illegible]

Pupil	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74
Spell.	36	40	40	43	30	41	46	53	37	29	31	35	30	38	[illegible]
Arith.	27	23	39	32	36	32	39	49	36	23	44	36	44	[illegible]	32
Synon.	43	30	29	37	32	17	62	48	50	31	29	27	40	54	40
Digit.	57	27	34	41	55	57	48	45	57	50	50	53	43	59	30
Provb.	35	31	31	24	35	48	68	56	40	31	31	24	45	50	40
D. Sen.	38	27	32	37	41	35	46	41	45	27	40	40	40	34	[illegible]
Rel'n.	44	28	25	31	12	55	44	44	33	25	[illegible]	38	36	51	[illegible]
Geom.	45	22	38	22	28	49	55	55	23	30	45	41	58	45	38
Fol. D.	32	21	38	29	35	46	48	48	38	35	41	29	43	51	43
Compl.	31	22	23	30	24	59	41	61	45	22	39	26	43	47	38
Sums	388	271	329	326	328	469	497	500	404	303	399	349	451	483	[illegible]
C. B.	99	67	73	90	85	112	145	120	94	61	[illegible]	[illegible]	[illegible]	[illegible]	93

Pupil	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89
Spell.	45	32	42	40	36	50	33	46	35	34	44	29	44	51	15
Arith.	41	27	27	32	36	46	36	32	44	39	34	36	49	49	34
Synon.	50	34	27	28	43	51	28	30	49	33	28	39	36	54	46
Digit.	34	24	64	64	45	57	39	39	64	55	22	22	43	45	50
Provb.	48	38	24	35	31	43	40	43	52	31	38	40	38	35	48
D. Sen.	45	25	28	30	43	54	32	45	45	38	23	32	40	45	44
Rel'n.	42	36	32	34	41	53	31	46	55	39	32	33	38	42	29
Geom.	61	52	25	38	33	64	38	41	45	26	43	45	52	41	37
Fol. D.	43	29	29	43	46	61	41	41	51	35	29	46	51	32	35
Compl.	34	37	33	37	48	66	35	39	42	33	39	33	43	42	36
Sums	443	334	331	381	402	545	353	402	482	363	332	355	434	436	374
C. B.	117	80	75	93	102	127	68	110	119	95	80	88	107	115	94

Pupil	90	91	92	94	95	96	101	102	103	104	105	107	108	109	110
Spell.	42	8	34	28	33	31	35	50	52	54	52	49	58	40	64
Arith.	41	36	34	32	32	51	62	62	54	44	49	44	49	49	51
Synon.	48	22	43	22	37	34	74	56	45	45	54	50	59	46	59
Digit.	60	30	50	29	39	73	50	60	55	24	69	29	50	62	48
Provb.	40	24	35	35	35	43	65	62	43	40	62	38	45	43	52
D. Sen.	45	41	51	19	45	32	47	40	48	40	41	56	53	45	47
Rel'n.	44	29	41	31	35	38	55	60	58	42	60	53	46	38	58
Geom.	43	28	32	24	49	41	45	66	55	31	45	45	52	38	49
Fol. D.	41	32	43	29	38	32	58	75	56	25	56	43	61	46	51
Compl.	49	23	38	26	32	33	44	71	59	45	45	39	72	58	50
Sums	453	273	401	275	375	408	535	602	525	390	533	446	545	465	529
C. B.	122	59	85	71	85	93	117	117	87	75	99	80	99	83	118

Pupil	111	112	114	115	118	119	120	121	122	123	124	125	127	128	129
Spell.	55	44	61	47	51	70	36	41	56	47	51	59	56	68	56
Arith.	51	44	54	39	51	56	56	46	41	44	51	58	79	49	49
Synon.	49	40	62	51	41	75	43	50	55	54	53	45	72	51	52
Digit.	62	45	69	50	27	75	57	55	41	32	53	45	57	68	60
Provb.	58	56	75	56	38	71	56	56	50	45	48	58	65	46	52
D. Sen.	40	45	49	56	51	55	33	41	61	45	48	26	56	47	48
Rel'n.	44	53	58	48	23	69	48	55	46	56	60	71	73	58	55
Geom.	45	38	81	49	41	77	55	43	47	52	47	69	81	61	49
Fol. D.	48	48	70	38	32	78	56	61	46	51	58	53	72	51	65
Compl.	45	37	69	39	44	72	54	52	56	45	48	48	61	52	58
Sums	497	450	648	473	399	698	494	500	499	471	517	532	672	551	544
C. B.	106	88	132	92	64	168	96	93	87	115	102	106	144	112	127

Pupil	130	131	132	133	134	135	136	138	139	140	141	142	143	144	145	146
Spell.	64	55	53	45	64	46	52	47	44	47	51	49	40	52	61	59
Arith.	49	64	69	60	41	39	51	54	58	46	39	51	62	67	51	56
Synon.	73	62	62	55	48	43	48	43	53	55	56	55	52	48	63	64
Digit.	43	50	34	64	60	53	69	34	75	62	48	53	30	53	48	62
Provb.	81	65	62	58	54	40	40	52	60	65	60	58	46	45	71	71
D. Sen.	54	58	52	46	49	38	49	45	41	65	44	43	46	49	52	57
Rel'n.	61	60	55	58	48	39	48	56	60	55	49	36	58	58	49	67
Geom.	72	69	52	52	49	38	58	66	69	35	40	41	49	66	52	49
Fol. D.	65	51	56	43	58	43	61	41	58	43	53	38	51	38	70	80
Compl.	63	63	63	55	53	44	44	46	60	42	53	51	48	41	70	60
Sums	625	597	558	536	524	423	520	484	578	515	493	475	482	517	587	625
C. B.	131	115	118	112	102	74	99	88	101	100	90	87	85	97	108	129

Appendix III.

Some Mathematical Reasoning with Regard to Criteria of Tests of Intelligence.

1. If we were to assume that each test measured only a general factor—one common to all the tests—and one or more factors specific to that test alone, then the relative degrees in which two tests correlate with the general factor, are expressed, subject to the

chance errors of the coefficients, by the relative degrees to which these two tests correlate with the other tests. This may be shown as follows. By formula for partial correlation,

$$r_{ac.i} = \frac{r_{ac} - r_{ai}\, r_{ci}}{\sqrt{(1 - r_{ai}^2)\,(1 - r_{ci}^2)}}$$

in which a and c are tests, i is a hypothetical, perfect measure of the general factor, r_{ac} is the coefficent of correlation between a and c, etc., and $r_{ac.i}$ is the coefficient of correlation between a and c which is due to factors other than the general one. But since by hypothesis, the general factor in the only source of correlation $r_{ac.i} = 0$.

Then $$r_{ac} = r_{ai}\, r_{ci}$$

and similarly, $$r_{bc} = r_{bi}\, r_{ci}$$

Therefore $$\frac{r_{ac}}{r_{bc}} = \frac{r_{ai}}{r_{bi}}$$

Similarly, $$\frac{r_{ai}}{r_{bi}} = \frac{r_{ad}}{r_{bd}} = \frac{r_{ae}}{r_{be}} = \frac{r_{af}}{r_{bf}} = \text{etc.}$$

An expression for the combined value of these ratios is given very approximately by the ratio of the average intercorrelation between a and the other tests to the average intercorrelation between b and the other tests.

2. Let us consider now a case in which there is no factor common to all the tests in the group. To take a very simple example, let us suppose we have four tests (nos. 1, 2, 3, and 4) testing abilities each of which is made up of five of the nine elements, A, B, C, D, E, F, G, H, and I, distributed as follows.

Test 1,	A	B	C	D	E				
Test 2,	A	B	C	D		F			
Test 3,			C	D	E	F	G		
Test 4,		B			E	F		H	I

Here it will be noted that no element is common to more than three abilities. Now the coefficient of correlation between two series of values is a measure of the percentage of elemental causes common to both.* And since the number of elemental causes common to

*For example, if five coins are tossed n times and each time the number of heads appearing is recorded, and if after each independent tossing, one coin is left lying, the other four tossed again, and the number of heads then appearing is recorded; then as n approaches infinity, the coefficient of correlation between the number of heads appearing by the independent tossing and the number of heads appearing by the dependent tossing approaches .20, attesting to the fact that one fifth of the causes affecting the number of heads in each throw (one coin in five) was common to both throws of a pair. If two of the five coins are left lying, the correlation will approach .40, if three are left lying, .60, if four, .80, and if five, of course, 1.00. Similarly for other numbers of coins and similarly for elements of abilities.

abilities, 1 and 2, is four out of five, the correlation between tests, 1 and 2, will tend to be .80. Three elements are common to abilities, 1 and 3, and 2 and 3. Therefore the correlations between tests 1 and 3 and between 2 and 3 will tend to be .60. And so on. The correlation table will therefore appear as follows.

Tests	1	2	3	4
1		.80	.60	.40
2	.80		.60	.40
3	.60	.60		.40
4	.40	.40	.40	
Sums	1.80	1.80	1.60	1.20

A table showing the number of elements common to each pair of abilities would appear as follows.

Tests	1	2	3	4
1		4	3	2
2	4		3	2
3	3	3		2
4	2	2	2	
Sums	9	9	8	6

It may be seen from this table that the number of times the elements of ability, 1, appear in the other three abilities is 9. The correlation spread of ability, 1, may therefore be said to be represented by the number 9. The number of times the elements of abilities, 2, 3, and 4, appear in the other three abilities are respectively 9, 8, and 6. We may say, then, that the relative values of the correlational spreads of the four tests are as 9 : 9 : 8 : 6. Now it may be seen that these are exactly the same proportions as 1.80 : 1.80 : 1.60 : 1.20, the sums of the coefficients in the first table. The latter values are equal respectively to the former values when each is divided by 5, the number of elements in each ability. Thus it may be seen that the sums of the correlations of each of the tests with all of the others afford meaures of the relative correlational spread of the tests.

3. The coefficient of correlation between any one weighted test and the weighted composite of a number of tests may be found

from the coefficients of intercorrelation and the weights by the formula (see Ref. 13) which may be stated in general as follows.

$$r_{wa(x_1b_1+x_2b_2+x_3b_3\ldots)} = \frac{\sum(x\sigma_b r_{ab})}{\sqrt{\sum x^2\sigma_b^2 + \sum xx\sigma_b\sigma_c r_{bc}}}$$

in which w, $x_{,1}$ $x_{,2}$ etc., are the weights given to the tests, a, b_1, b_2, etc., and σ_b is the standard deviation of the scores of any test, b.

McCall's procedure, therefore, might have been to consider b_1, b_2, etc., as representing the tests which he wished to embody in his Composite; to consider x_1, x_2, etc., as representing the respective weights to be given these tests, and a as representing any test it was desired to correlate with the Composite. The correlation could then have been obtained by solving the equation. The general formula, is equally applicable, of course, for finding, from the intercorrelations, the correlation of a test with the average of all the *other* tests.

If only the relative values of the correlations of each test with a composite of weighted tests is desired, these may be obtained more simply yet; thus, assuming that there were only three tests, a, b, and c, in the group, weighted respectively, w, x, and y, then the corresponding formula for the correlation of test a, with the weighted composite would be

$$r_{wa(wa+xb+yc)} = \frac{w\Sigma_a r_{aa} + x\Sigma_b r_{ab} + y\Sigma_c r_{ac}}{\sqrt{w^2\sigma_a^2 + x^2\sigma_b^2 + y^2\sigma_c^2 + 2(w\sigma_a x\sigma_b r_{ab} + w\sigma_a y\sigma_c r_{ac} + x\Sigma_b y\Sigma_c r_{bc})}}$$

Similarly, $$r_{xb(wa+xb+yc)} = \frac{w\Sigma_a r_{ab} + x\Sigma_b r_{bb} + y\Sigma_c r_{bc}}{\text{same denominator}}$$

And $$r_{yc(wa+xb+yc)} = \frac{w\Sigma_a r_{ac} + x\Sigma_b r_{bc} + y\Sigma_c r_{cc}}{\text{same denominator}}$$

Since the denominators are the same in all cases, it may be seen that the relative values of the correlations of the several tests with the weighted composite are directly proportional to the sums of the intercorrelations of those tests, each with all the tests, when these intercorrelations have been weighted as shown in the numerators. And if the standard deviations have been made equal, this merely means weighting the several coefficients by the same weights in and the same order as they would appear in the composite. The same reasoning holds for any number of tests.

4. If it is desired, on the other hand, to find the relative or absolute amounts of the average intercorrelations of each of a series with all the others, (weights and standard deviations being equal) as a criterion of the degree to which each test measures the common factor; and if the values of the separate intercorrelations are not

required; it will be more convenient to derive these average intercorrelations from the correlation of each test with the average of all the measures taken together as a composite. That this may be done is shown as follows.

Repeating the proof given in 3 above in a simpler form, let us assume again for the moment that there are only three tests, *a*, *b*, and *c*, in the series; then by the formula for the correlation of one test with the average of a number of tests (assuming weights and standard deviations equal),

$$(1) \qquad r_{a\,(a+b+c)} = \frac{r_{aa}+r_{ab}+r_{ac}}{\sqrt{3+2(r_{ab}+r_{ac}+r_{bc})}}$$

$$(2) \qquad r_{b(a+b+c)} = \frac{r_{ba}+r_{bb}+r_{bc}}{\sqrt{3+2(r_{ab}+r_{ac}+r_{bc})}}$$

$$(3) \qquad r_{c(a+b+c)} = \frac{r_{ca}+r_{cb}+r_{cc}}{\sqrt{3+2(r_{ab}+r_{ac}+r_{bc})}}$$

Letting $\Sigma r_{a,b,c(a+b+c)}$ represent $r_{a(a+b+c)}+r_{b(a+b+c)}+r_{c(a+b+c)}$, and since $r_{aa}=r_{bb}=r_{cc}=0$,

$$(4) \qquad \Sigma r_{a,b,c(a+b+c)} = \frac{3+2(r_{ab}+r_{ac}+r_{bc})}{\sqrt{3+2(r_{ab}+r_{ac}+r_{bc})}}$$

$$(5) \qquad \Sigma r_{a,b,c\,(a+b+c)} = \sqrt{3+2(r_{ab}+r_{ac}+r_{bc})}$$

Multiplying equation 1 by equation 5, we have

$$r_{aa}+r_{ab}+r_{ac} = r_{a(a+b+c)} \times \Sigma r_{a,b,c\,(a+b+c)}$$

and similarly for all other correlational sums. Thus it may be seen that the *absolute* amounts of the sums or averages of the intercorrelations of any test with all the tests in the series may be derived from the values of the correlations of each test with the composite (weights and standard deviations being equal) without the individual test intercorrelations being found. The same reasoning holds for any number of tests. As a criteria of the degree to which any test measures the factor common to the group of abilities tested, the test's average intercorrelation with all the tests and its correlation with the composite are of equal value, being, in fact, the same criterion. The sums of the intercorrelations of any test with all the *other* tests (excluding itself) may be obtained, of course, by subtracting 1.00 (the correlation of the test with itself) from the sum of the intercorrelations with *all* the tests.

APPENDIX IV

Inter-test Correlations (Raw)

	Synonyms	Narrative Completion	Following Directions	Proverbs	Relation	Arithmetic	Spelling	Geometric	Disarranged Sentences	Digit
Synonyms		.798	.781	.828	.750	.722	.753	.674	.699	.322
Narrative Completion	.798		.815	.738	.742	.719	.787	.689	.654	.341
Following Directions	.781	.815		.762	.749	.686	.662	.716	.561	.420
Proverbs	.828	.738	.762		.738	.700	.686	.677	.567	.358
Relation	.750	.742	.749	.738		.746	.688	.719	.612	.282
Arithmetic	.722	.719	.686	.700	.746		.662	.687	.705	.324
Spelling	.753	.787	.662	.686	.688	.662		.632	.679	.259
Geometric	.674	.689	.716	.677	.719	.687	.632		.570	.301
Disarranged Sentences	.699	.654	.561	.567	.612	.705	.679	.570		.280
Digits	.322	.341	.420	.358	.282	.324	.259	.301	.280	
Average	.703	.698	.684	.673	.670	.661	.645	.630	.592	.321

Inter-test Correlations (Corrected for Attenuation)

	Synonyms	Narrative Completion	Relation	Proverbs	Following Directions	Arithmetic	Geometric	Spelling	Disarranged Sentences	Digit
Synonyms		.901	.881	.960	.887	.808	.769	.825	.819	.378
Narrative Completion	.901		.845	.830	.898	.779	.762	.836	.742	.386
Relation	.881	.845		.864	.858	.842	.828	.760	.723	.332
Proverbs	.960	.830	.864		.862	.780	.770	.749	.662	.416
Following Directions	.887	.898	.858	.862		.749	.798	.708	.642	.478
Arithmetic	.808	.779	.842	.780	.749		.753	.697	.793	.363
Geometric	.769	.762	.828	.770	.798	.753		.679	.655	.344
Spelling	.825	.836	.760	.749	.708	.697	.679		.748	.280
Disarranged Sentences	.819	.742	.723	.662	.642	.793	.655	.748		.329
Digits	.378	.386	.332	.416	.478	.363	.344	.280	.329	
Averages	.803	.775	.770	.766	.761	.729	.706	.698	.679	.36[illegible]

REFERENCES

1. A. R. Abelson. *Mental Ability of Backward Children.* Brit. Jr. of Psy., Dec., 1911, 268–314.

2. Leonard P. Ayres. *Measurement of Ability in Spelling.* Russell Sage Foundation, Educational Monograph E139.

3. Cyril Burt. *Experimental Tests of General Intelligence.* Brit. Jr. Psy., Vol. iii, 94.

4. F. J. Kelly. *The Kansas Silent Reading Tests.* Bureau of Educational Measurements and Standards, Emporia, Kansas.

5. Wm. A. McCall. *Correlation of Some Educational and Psychological Measurements.* Columbia University Contributions to Education, No. 79.

6. M. Merriman. *Method of Least Squares.* John Wiley and Sons, N. Y., 1894, Ch. IV.

7. Arthur S. Otis. *The Reliability of Spelling Scales.* School and Society, Vol. IV, Nos. 96, 97, 98, and 99.

8. Arthur S. Otis. *Considerations Concerning the Making of a Scale for the Measurement of Reading Ability.* Ped. Sem., Dec., 1916, Vol. XXIII, pp. 528–549.

9. Arthur S. Otis. *Some Logical Aspects of the Binet Scale.* Psy. Rev., Vol. XXIII, Nos. 2 and 3.

10. Arthur S. Otis. *A Criticism of the Yerkes-Bridges Point Scale, with Alternative Suggestions.* Jr. Ed. Psy., Mar., 1917.

11. Arthur S. Otis. *The Reliability of the Binet Scale and Pedagogical Scales.* (To be published).

12. Daniel Starch. *Educational Measurements.* The Macmillan Company, 1916.

13. C. Spearman. *Correlations of Sums and Differences.* Brit. Jr. Psy., Vol. 5, 419–426.

14. Lewis M. Terman. *The Measurement of Intelligence.* Houghton Mifflin Co., 1916.

15. M. R. Trabue. *Completion Test Language Scales.* Col. Univ. Cont. to Ed., No. 77, 1916.

University of Toronto
Library

DO NOT
REMOVE
THE
CARD
FROM
THIS
POCKET

Acme Library Card Pocket
Under Pat. "Ref. Index File"
Made by LIBRARY BUREAU

155661
Psych.
O88
Author Otis, Arthur Stinton
Title An absolute point scale for the group measurement